Real-Time Visual Computations Using Analog CMOS Processing Arrays

Massimo A. Sivilotti, Michelle A. Mahowald and Carver A. Mead
California Institute of Technology
Pasadena, California 91125

Abstract

Integration of photosensors and processing elements provides a mechanism to concurrently perform computations previously intractable in real-time. We have used this approach to model biological early vision processes. A set of VLSI "retina" chips have been fabricated, using large scale analog circuits (over 100K transistors in total). Analog processing provides sophisticated, compact functional elements, and avoids some of the aliasing problems encountered in conventional sampled-data artificial vision systems.

Integration of Photoreception and Processing

By their very two-dimensional nature, images constitute a high bandwidth interface with the real world. The most powerful supercomputers are incapable of even rudimentary analysis of static images. Real-time analysis of motion information, requiring computation over several images, is completely infeasible. Yet biological early vision processes are clearly able to perform these computations, by exploiting the inherent parallelism of visual inputs in a truly concurrent fashion. Computations are spatially localized, and computing elements are replicated as required. Only significant information is transmitted along the optic nerve.

Artificial vision systems are further limited by the early sampling performed by television camera front-ends. Because each point is sampled only every 1/30 second, an object can easily move several pixels be-

tween samples. Motion interpretation has thus been converted from a local problem to the much more difficult *correspondence problem.* In signal processing terms, high frequency information is irretrievably lost due to *aliasing.*

Modeled on biological architectures, our approach is to spatially interleave integrated photoreceptors and processing on a VLSI die [16]. To obtain a sufficiently rich, yet compact, set of computing elements, analog micropower CMOS circuits are used. The resulting high density permits complex systems to be built, that demonstrate powerful collective behaviors [20]. Finally, by performing temporal operations on continuous data, prior to sampling for transmission off-chip, susceptibility to aliasing is reduced.

Models for Retinal Computations

The retina performs the first step in visual processing and provides the data for all subsequent stages of the visual system. Although various species perform slightly different sets of retinal computations, there are several aspects of visual processing that are common to many different organisms [6]. These ubiquitous features include logarithmic compression of the incoming intensity at the detector level, and the extensive use of lateral and temporal inhibition in the retinal computations. These functions are computed using smoothly varying (continuous) analog potentials, rather than neuronal action potentials.

Several explainations have been proposed for why the visual system performs these computations in the retina. As is often the case when investigating biological systems, it is not possible to determine *the* reason that the system adopted a particular strategy; rather, these systems are optimized with respect to multiple constraints [18].

One particular set of advantages to retinal processing can be observed by assuming that the function of the retina is the "neat packaging of information" [2] to be sent to higher visual areas. Retinal ganglion cells transmit information to the brain by propagating action potentials along their axons in the optic nerve. There are a finite number of discriminable signal levels coming over the optic nerve due to intrinsic noise. The visual information must be encoded in such a way that

the full dynamic range of the neuron is used. Automatic gain control mechanisms, such as logarithmic compression of intensity and center-surround antagonistic receptive fields allow the system to encode detail over a large range of ambient light levels. In addition to the limited resolution of signal levels, the discrete-time nature of the action potential limits the temporal resolution of events. A large fraction of retinal processing is dedicated to extracting motion events. If retinal ganglion cells encoded simple intensity, then any change in intensity would be encoded as a change in pulse rate. Even assuming that such an encoding allowed no statistical fluctuation in pulse interval, the time at which such a change had occurred could be determined only to the time between pulses. In signal processing terms, the derivative information would have been aliased away by temporal sampling of the image. For this reason, motion detecting ganglion cells produce action potentials that correspond to changes in intensity, rather than intensity itself. In this way, a pulse burst corresponds to an important feature moving over that particular place on the retina. Higher–level correlations among events can then be reconstructed without loss of information due to temporal aliasing.

Light-Level Independence

A vision system intended for operation in an unconstrained environment must include automatic gain control (AGC) with respect to absolute ambient light level. Taking the logarithm of the incident light intensity is a simple local AGC mechanism. Receptors with logarithmic response have the additional advantage of providing output voltages whose difference is proportional to the *contrast ratios* within the image, which are the perceptually important parameters.

An integrated photoreceptor with an output that is logarithmic over 5 orders of magnitude in light intensity is shown in Figure 1a. Its operation is similar to that of one previously described [9]; a large-area bipolar transistor is formed using the *n*-well for the base and *p*-type diffusion as the emitter. The substrate forms the collector, and hence the device is operated in a common-collector configuration. The output voltage biases the gate of a *p*-channel MOS feedback transistor operating in subthreshold. In this regime, the channel current is exponential in the gate voltage with a slope of about 1 decade per 100 mV.

If a second subthreshold transistor is used for source degeneration, the slope can be decreased to about 1 decade per 300 mV. This arrangement provides a larger output voltage swing, and sets the output voltage in the 1.0 to 2.5 V range below V_{dd}, making direct coupling to subsequent stages possible. The feedback current is generated by mirroring the load current for the emitter of the phototransistor. If the feedback to the phototransistor base is omitted (Figure 1b), the receptor is sensitive to lower light levels (by a factor of h_{fe}), but will saturate at bright levels, as the MOS loads leave subthreshold. These receptors operate at light levels comparable to the useful range of cones in human retinas, and form the basis for the RET10 chip discussed later.

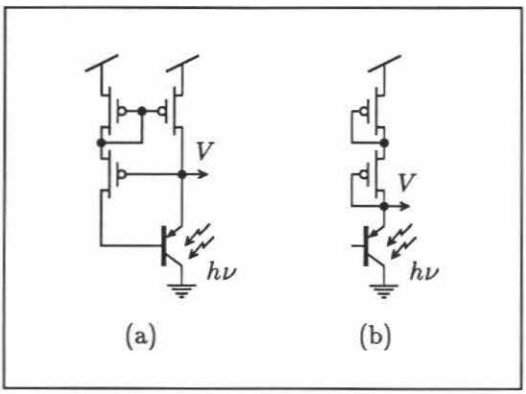

Figure 1: Logarithmic photoreceptors.

A Simple Local Computation: Discrete-Time Derivative

We perceive motion when a point in an image displays non-zero spatial and temporal derivatives. In other words, an edge (space derivative) that is moving causes a change in brightness at that point in the image. Thus, a local time derivative is the simplest computation to highlight areas of an image that are moving. Because this computation is purely local, no interpixel communication is required. The derivative can be approximated by comparing the present photoreceptor output with some suitably delayed version of the output.

A discrete-time derivative based on the circuit in Figure 2 forms the core of the RET20 chip. When switch S_1 is closed, the amplifier forms

a unity-gain follower stage that stores the current state of the system on capacitor C_1. The switch is then opened, and any evolution of the input away from the sampled value is amplified by the open-loop voltage gain of the (wide-range) transconductance amplifier [21].

If switch S_1 is implemented with a MOS pass transistor, transient switching charge is injected onto C_1 [15]. To minimize this effect, a transconductance amplifier was used for the switch (Figure 3). When the bias current in A_2 is reduced to zero, the capacitor is effectively isolated from the input. Less noise charge is generated because the channel charge is symmetrically divided between both branches of the differential pair (which have identical operating points in the follower configuration), and because any capacitive clock feedthrough is decreased by the cascode connection of the differential pair. In addition, the clock need not be rail-to-rail (hence decreasing $\frac{dV}{dt}$ even further), and it can be single phase and unipolar. Figure 4 illustrates the output of the circuit when presented with an asymmetrical 100 Hz triangle wave input (ϕ clock frequency 1000 Hz).

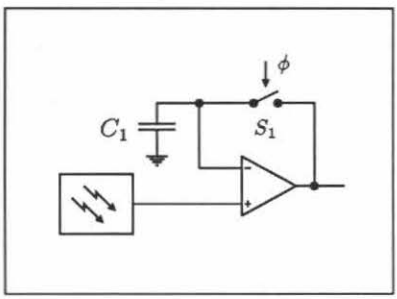

Figure 2: Photoreceptor with discrete-time differentiator.

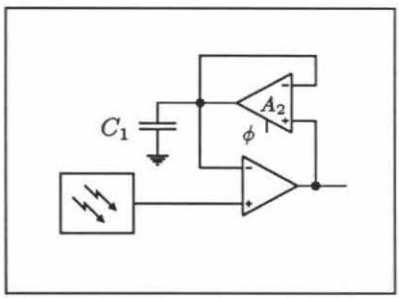

Figure 3: Transconductance amplifier used as low-noise switch.

RET30: Continuous-Time Derivatives and Local Space Derivatives

Biological retinas contain horizontal cells that provide lateral conductance and can be loosely thought of as providing an average of the signal values in the neighborhood with which the local signal can be compared. Inspired by this model, the RET30 (Figure 5) consists of

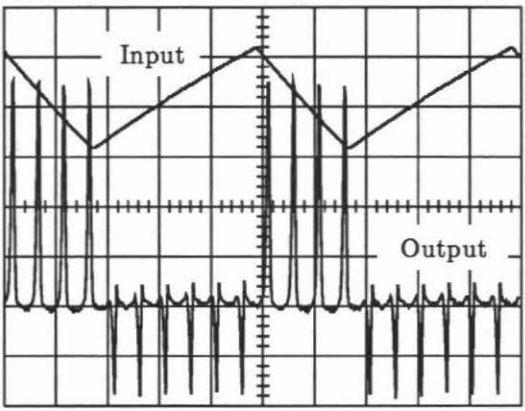

Figure 4: Discrete-time derivative: sample operation.

an array of receptors, R, interconnected by a hexagonal resistive network [11]. To provide temporal smoothing, a capacitor to ground is located at each junction of 6 neighboring horizontal resistors.

Each local processing element takes the difference between the potential of the horizontal network and that of the receptor output, and drives the local potential of the horizontal network toward the local receptor output potential. The "derivative" computed is the difference between the input signal and a spatially and temporally smoothed version of that signal. The spatial part of the processing emphasizes areas in the image containing the most information. The emphasis corresponds to a discrete approximation to a Laplacian operator applied to the image. The temporal part of the processing is a finite-gain, single time-constant differentiation.

Horizontal resistors: To construct a practical space–time derivative system, we must be able to create time constants of the same order as the time scale of motion events, without using enormous area for capacitors. The horizontal network spreads the potential at one point outward through a resistive sheet. To keep the time constant (τ_h) of the spreading on the same scale as others in the system, enormous resistor values are required, ($10^{11} - 10^{13}\Omega$) — larger than the resisitance of any integrated device we can build. A circuit that implements a very high-value resistance in a controlled way is shown in Figure 6.

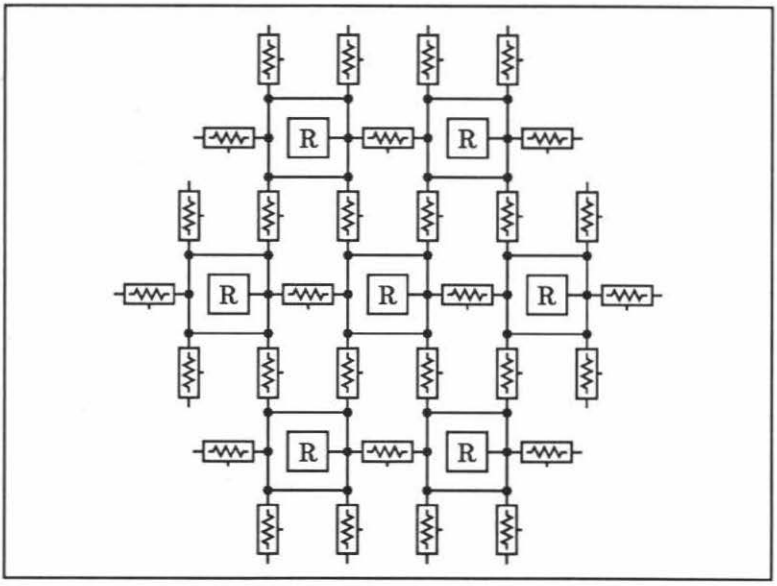

Figure 5: RET30 tesselation.

V_1 and V_2 represent potentials V_i of two neighboring locations in the network. The current I_0 into the upper node is constrained to be the same as that out of the lower node by a current-mirror arrangement. Hence, any current out of node 1 must flow into node 2. By symmetry, the magnitude of this current must be zero when the voltages are equal, and will be monotonically related to their difference. However, it can never exceed I_0. The I–V relationship is shown in Figure 7. The limiting current I_0 is set by the current mirror input, and controls the value of τ_h. As can be seen from the figure, the region of linear operation of the horizontal circuits is about ± 100 mV, corresponding to an illumination contrast ratio of about 2:1 at the photoreceptors.

Time differentiators: The differentiator of the RET30 chip is implemented using the same circuit as that shown in Figure 3, except the A_2 amplifier is not clocked. Because the ϕ input on the second amplifier controls the maximum current that may flow into or out of the storage capacitor, it determines the rate at which the capacitor is charged, and hence the time-constant τ_r. The net current into the capacitor is of the same form as that shown in Figure 7, for the same

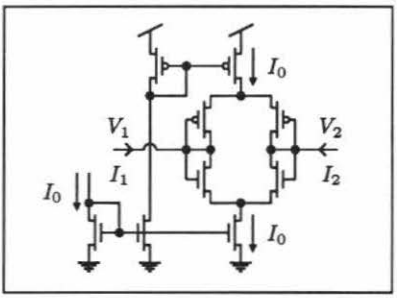

Figure 6: Horizontal resistor.

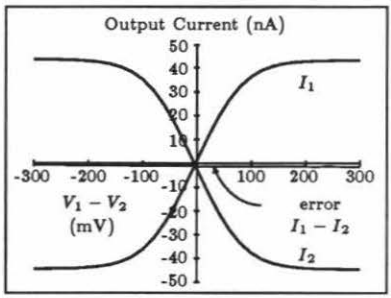

Figure 7: I–V characteristic.

reasons. This circuit has unity gain at DC, and a gain at short times set by the open-circuit gain of the first amplifier, as in the discrete-time case. Experimental data illustrating the operation of the continuous-time differentiator are shown in Figure 8.

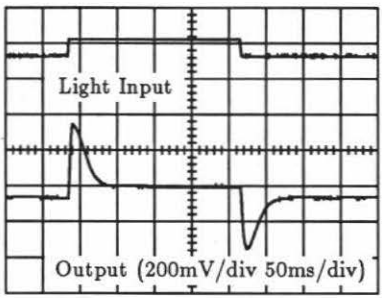

Figure 8: Continuous-time differentiator.

The advantage of this arrangement is that any input offset voltage of the first amplifier is not multiplied by the gain of the amplifier in its effect on the output voltage.

It can be argued that the saturating characteristic of all these circuits is desirable, as it prevents one extreme input (or faulty circuit element) from paralyzing the entire network. Thus, even for these simple operations, many of the properties of collective circuits can be preserved.

As shown, maximum outputs will occur when high contrast features move over the retina. If only time derivative information is desired, the horizontal network is unnecessary, and can be disabled by setting the I_0

input current to zero. When enabled, the horizontal network computes the extent to which the light received by an individual receptor differs from the average level in its neighborhood. It is thus most sensitive to a point, less to a corner, even less to an edge, and not at all to a uniform gradient. The system can be made to display a sustained response to one of its preferred stimuli even if that stimulus is stationary.

CMOS Design Frame for Scanning Arrays

Fundamental limits on the number of pads imposed by available VLSI packaging technologies, coupled with the high area cost of dedicated wiring within the imaging array, necessitate a scalable communication architecture. Time multiplexing of sampled data signals requires a minimal number of pads, and simplifies the external system design by reducing the number of external components needed to compensate for the different electrical nature of the off-chip environment.

To facilitate experimentation with a wide variety of processing core cells, a "design frame" approach was adopted, with a standard peripheral frame providing all the communication support functions, as well as the data sampling mechanism. By separating the computation and communication tasks, the responsibility of the core cell designer is simplified, and consists of providing, as the result of some computation, an analog voltage to be sampled and transmitted. An additional benefit is the independence of the tiling topology from the computation topology. For example, our design frames support true hexagonal tiling (Figure 9), hexagonal tiling using offset rectangles (Figure 10), and pure rectangular tiling; in all cases, the pixel stream is (offset) rectangular.

Considerable work has been done on charge-transfer systems, particularly for CCD image arrays [5]. Recently, bipolar phototransducer arrays have received some attention due to their more favorable saturation and antiblooming properties [7,1,3]. In general, the non-CCD scanning techniques consist of switching some charge basin onto a column line, then switching that column's charge packet onto a global output line. Often, these switches are implemented with specially fabricated low-threshold MOS devices [14]. The primary source of fixed-pattern noise on the output is due to inversion charge variation

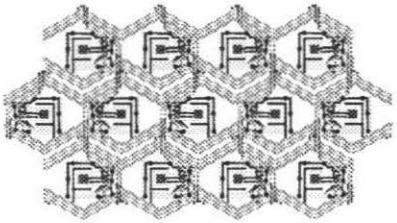

Figure 9: Hexagonal tiling with Boston geometry (HEXRET).

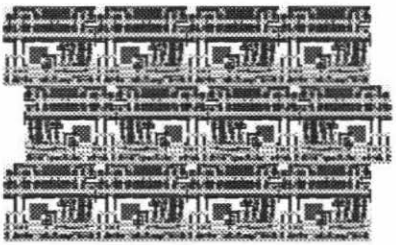

Figure 10: Hexagonal tiling with Manhattan geometry (RET30).

in the switching transistors [13,23]. Mechanisms to reduce this noise include integration over the entire pixel time, and sampling at some constant point during the pixel event.

Our RETxx chips have taken a somewhat different approach to obtaining an acceptable signal-to-noise ratio (SNR). First, photocharge is integrated at each pixel site. It is not then destructively dumped on the output line, but rather is stored on a local capacitor, which is nondestructively sampled using a single MOS charge–sense transistor. The *current* in this "bit line" is sensed, eventually by an external amplifier. To minimize signal propagation delays due to $C\frac{dV}{dt}$ losses (*i.e.* charging/discharging the highly capacitive bit lines, output line, and output pad/off-chip wiring), current-steering sensing is employed throughout.

Pixels are enabled on a row-by-row basis by a switched pass transistor in series with the transduction transistor (Figure 11). Two configurations are possible: Figure 11a shows a conventional cascode arrangement that minimizes the dependence of I on V_{bus}; Figure 11b shows a configuration designed to maximize the linear range of I with V, by operating M_2 in the ohmic region to provide source degeneration for M_1. In either case, linear operation can be guaranteed if V_{bus} is maintained sufficiently close to ground, so M_1 is operating in its ohmic regime.

One entire row of the scanning array is enabled simultaneously, at the line clock rate ϕ_V. Within each scan line, the pixel clock ϕ_H sequences the connection of the different "bit" lines onto the single

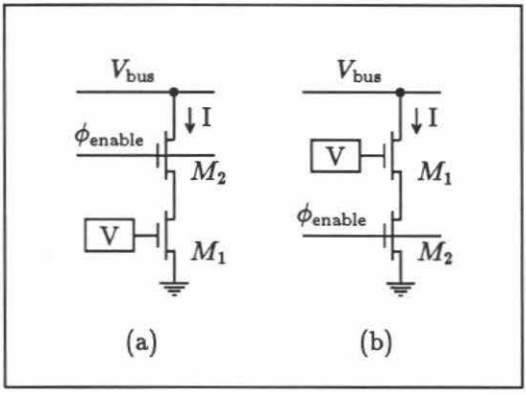

Figure 11: Current transducers.

output line (Figure 12). To maintain control of V_{bus}, and to keep this value independent of the pixel current I, the $N-1$ bit lines not connected to the output bus are instead connected to a dummy bus, biased at the desired value of V_{bus} (Figure 13). Thus, the pixel current I flows at all times.

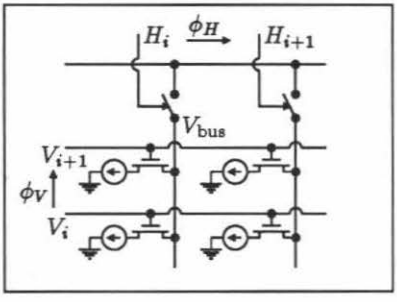

Figure 12: Horizontal switches.

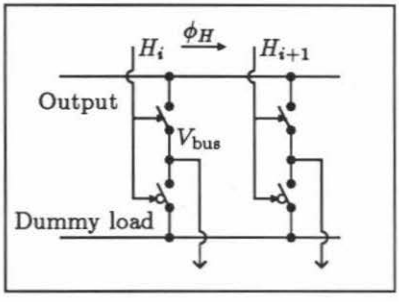

Figure 13: Dummy load line.

Off-chip, the current in the output line is converted to a voltage by a current–sense amplifier (Figure 14). This configuration implicitly biases the output line at V_{bus}. External compensation in the form of C_{comp} is required to counteract the highly capacitive input seen by the opamp.

The last circuit elements to consider in the analog signal path are the multiplexing switches in Figure 13. These can inject noise charge onto the output lines in two ways: (1) as a MOS transistor shuts

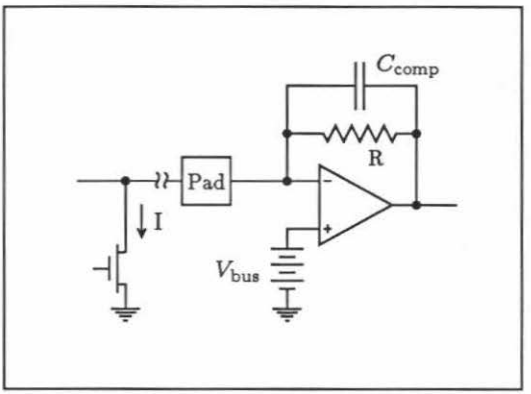

Figure 14: External current–sense amplifier.

off, the mobile charge in the channel region is divided (nearly equally) between the source and the drain, and (2) the switching clock itself can couple through the overlap capacitance of the gate with the source–drain regions. To minimize these problems, CMOS transmission gates driven by complementary phases derived from the horizontal clock were used as switches. Any injected parasitic charge is offset by the opposite physical process occurring in the complementary transistor, as well as by the same process occurring in reverse in an adjacent pass gate (as that column becomes enabled). More complicated charge compensation schemes are possible [8,19], but our simulation results indicated that these schemes did not significantly improve the clock noise suppression in this case.

Compensation of charge injection using complementary devices depends on the availability of a clock and its logical complement with *minimal skew* between these signals. In the presence of appreciable skew, the transmission gate switch actually can be noisier than a single pass transistor switch. For this reason, the clock signal and its complement are generated simultaneously by a CMOS set–reset logic (CSRL) [12] shift register, in which both signals are propagated together. This shift register is clocked with a two-phase nonoverlapping clock running at the pixel rate. Because the CSRL outputs are not fully restored during one clock phase, they are buffered by a pair of inverters. The full circuit, including multiplexing switches, is shown in Figure 15. The input to the shift register chain is brought off-chip. During one line period, a single "1" is shifted into the register, and is

shifted sequentially through all the pixel columns.

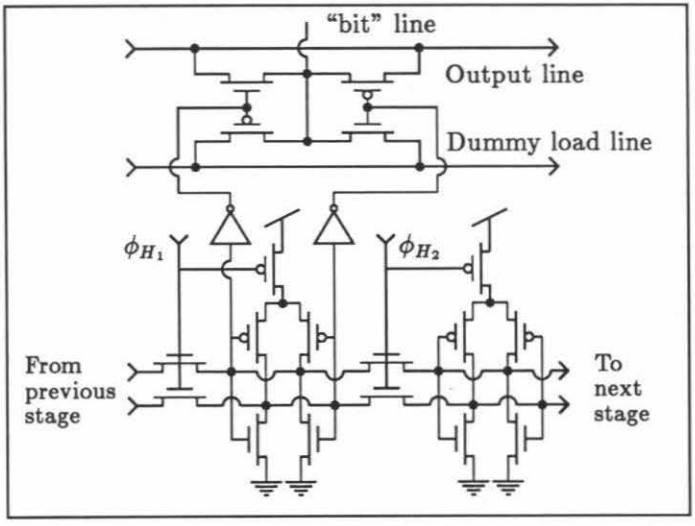

Figure 15: CSRL stage with multiplexer.

The same CSRL design is used in the vertical shift register to sequentially enable successive rows. The buffer inverters serve the additional function of driving the highly capacitive "word line" (ϕ_{enable} in Figure 11).

A functional diagram of the entire system is shown in Figure 16. An objective of the design frame was to simplify the system-level interface with the pixel array. There are only eight signal lines (four clocks — ϕ_{V_1}, ϕ_{V_2}, ϕ_{H_1}, ϕ_{H_2}; two shift register inputs — H_{in}, V_{in}; and two analog outputs — I_{out}, $I_{\text{dummyload}}$) plus power and any signals specific to the core cells in the pixel array.

Implementation and Experimental Results

We fabricated three different chips, incorporating the RET10, RET20, and RET30 core cells. Pixel array sizes and chip dimensions are shown in Table 1. The first versions were fabricated on MOSIS [4] run M57Q in August 1985. Except for the HEXRET, they all share the same design frame, implemented in Manhattan geometry using the WOL

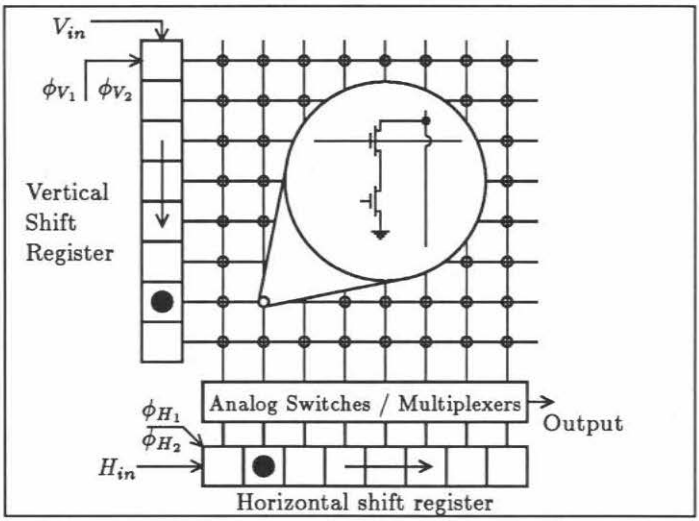

Figure 16: Design frame: system level interface.

design tool set [10,17]; the HEXRET implemented the RET10 circuit using arbitrary angle geometry, and was layed out with the TIGGER/POOH design system [22].

Table 1: Performance of Core Cells (3 μm feature size)

Chip	Pixel size ($\mu m \times \mu m$)	Array size	Measured SNR (dB)
phototransistor	33×33	N/A	
RET10	92×80	88×88	62
RET20	113×98	64×64	70
RET30	164×143	48×48	60

The chips were tested with two different setups: (1) a workstation-based tester capable of running all the chip clocks, and of digitizing and displaying the output signal at a rate of 60,000 pixels/sec (approximately 8 frames/sec for RET10), and (2) a TTL clock generator board that also produced a raster on which the chip output could be superposed and displayed on an oscilloscope at rates of more than 400,000 pixels/sec (over 50 frames/sec).

Figure 17 shows sample output from the RET10 chip, and clearly illustrates the logarithmic compression performed by the photoreceptors.

Figures 18 and 19 demonstrate the operation of the RET30 chip. In both cases, the stimulus is a dark cross mounted on a rotating axis. In Figure 18 the cross is stationary and, with the horizontal resistor network disabled, no image is seen. In Figure 19 the cross is rotating at approximately 10 rpm, and is clearly visible.

Figure 17: RET10 sample output.

Conclusions

Although many models have been proposed for the visual system, it is not possible to simulate enough cases to gain real confidence in the model, even on our most powerful computers. For this reason, we will not really understand visual processing, especially with respect to demanding problems such as motion analysis, until we succeed in building a system that is capable of doing visual processing in real-time. Until recently, we have not had a technology in which such fundamental synthetic investigations could be carried out. With the evolution of high-density VLSI technology, we have a means for conducting these extremely important investigations. The work will not be trivial. Previously, the most massive application of large-scale circuits has been in digital systems. Although analog integrated circuit techniques have developed along with digital ones, no comparable methods exist for managing the complexity of extremely large analog systems. This paper not only has described a prototype vision system, but also has illustrated an approach to problems of this class.

Acknowledgements

This research was supported by a grant from the System Development Foundation, and an equipment grant from Hewlett-Packard Corp. The authors are indebted to Dick Lyon, Lars Neilsen, Michael Emerling, and John Tanner for many useful discussions.

References

[1] M. Aoki, H. Ando, S. Ohba, I. Takemoto, S. Nagahara, T. Nakano, M. Kubo, and T. Fujita. 2/3-inch format MOS single-chip color imager. *IEEE Journal of Solid State Circuits*, SC-17(2):375–380, April 1982.

[2] H. B. Barlow. Three points about lateral inhibition. In W.A. Rosenblith, editor, *Sensory Communication*, pages 782–786, M.I.T. Press, Cambridge Mass., 1961.

[3] Savvas G. Chamberlain and Jim P.Y. Lee. A novel wide dynamic range silicon photodetector and linear imaging array. *IEEE Journal of Solid State Circuits*, SC-19(1):41–48, February 1984.

[4] D. Cohen and G. Lewicki. MOSIS—the ARPA silicon broker. In *Proceedings from the Second Caltech Conference on VLSI*, pages 29–44, California Institute of Technology, Pasadena, CA, 1981.

[5] P.L.P. Dillon, D.M. Lewis, and F.G. Kasper. Color imaging system using a single CCD area array. *IEEE Trans. Electron Devices*, ED-25:102–107, February 1978.

[6] Akimichi Kaneko. Physiology of the retina. *Ann. Rev. Neurosci*, 2:169–191, 1979.

[7] N. Koike, I. Takemoto, K. Satoh, S. Hanamura, S. Nagahara, and M. Kubo. MOS area sensor: part I—design consideration and performance of an *n-p-n* structure 484 × 384 element color MOS imager. *IEEE Journal of Solid State Circuits*, SC-15(4):741–746, August 1980.

[8] J.L. McCreary and P.R. Gray. All-MOS charge redistribution analog-to-digital conversion techniques—part I. *IEEE Journal of Solid State Circuits*, SC-10:371–379, December 1975.

[9] Carver Mead. A sensitive electronic photoreceptor. In *1985 Chapel Hill Conference on Very Large Scale Integration*, pages 463–471, 1985.

[10] Carver Mead. *The WOLery*. Technical Report 5113:TR:84, California Institute of Technology, June 1984.

[11] Carver Mead and Michelle Mahowald. An electronic model of the Y-system of mammalian retina. California Institute of Technology, Internal Technical Memo 5144:DF:84, June 1984.

[12] Carver Mead and John Wawrzynek. A new discipline for CMOS design: an architecture for sound synthesis. In *1985 Chapel Hill Conference on Very Large Scale Integration*, pages 87–104, 1985.

[13] S. Ohba, M. Nakai, H. Ando, S. Hanamura, S. Shimada, K. Satoh, K. Takahashi, M. Kubo, and T. Fujita. MOS area sensor: part II—low-noise MOS area sensor with antiblooming photodiodes. *IEEE Journal of Solid State Circuits*, SC-15(4):747–752, August 1980.

[14] Yoshio Ohkubo. An analysis of fixed pattern noise for MOS-CCD type image sensors under quasi-stationary conditions. *IEEE Journal of Solid State Circuits*, SC-21(4):555–560, August 1986.

[15] Bing J. Sheu and Chenming Hu. Switch-induced error voltage on a switched capacitor. *IEEE Journal of Solid State Circuits*, SC-19(4):519–525, August 1984.

[16] Massimo A. Sivilotti. *Toward a Motion-Based VLSI Vision System.* Master's thesis, California Institute of Technology, 1986. 5225:TR:86.

[17] Massimo A. Sivilotti. *A User's Guide to the WOL Design Tools.* Technical Report 5237:TR:86, California Institute of Technology, 1986.

[18] M. V. Srivivasan, S. B. Laughlin, and A. Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. London B*, 216:427–459, 1982.

[19] R.E. Suarez, P.R. Gray, and D.A. Hodges. All-MOS charge redistribution analog-to-digital conversion techniques–part II. *IEEE Journal of Solid State Circuits*, SC-10:379–385, December 1975.

[20] John Edward Tanner. *Integrated Optical Motion Detection.* PhD thesis, California Institute of Technology, 1986. 5223:TR:86.

[21] Eric A. Vittoz. Micropower techniques. In Yannis Tsividis and Paolo Antognetti, editors, *Design of MOS VLSI Circuits for Telecommunications*, pages 104–144, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[22] Telle E. Whitney. *Hierarchical Composition of VLSI Circuits.* PhD thesis, California Institute of Technology, 1985. 5189:TR:85.

[23] W.B. Wilson, H.Z. Massoud, E.J. Swanson, R.T. George Jr., and R.B. Fair. Measurement and modeling of charge feedthrough in n-channel MOS analog switches. *IEEE Journal of Solid State Circuits*, SC-20(6):1206–1213, December 1985.

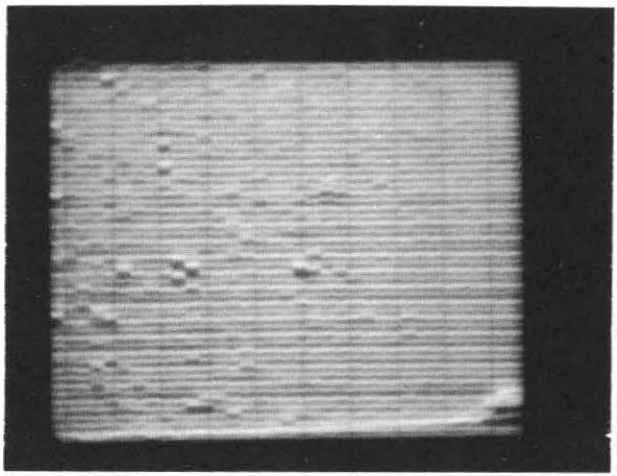

Figure 18: RET30 — stationary image.

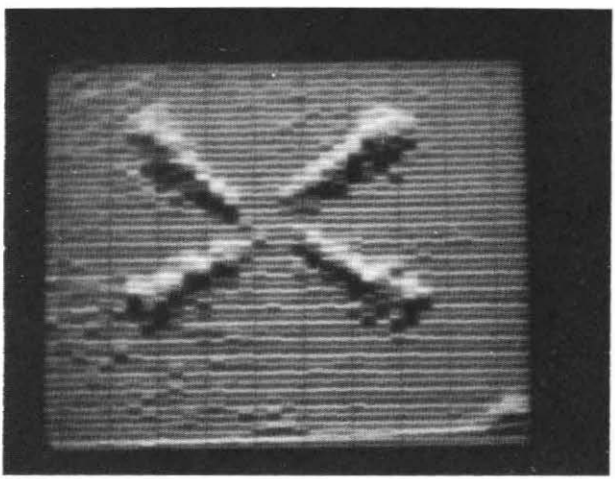

Figure 19: RET30 — moving image.